Tutorial

Research Synthesis Methods

Received 21 November 2014, Revised 15 May 2015, Accepted 16 May 2015 Published online 19 August 2015 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jrsm.1160

Get real in individual participant data (IPD) meta-analysis: a review of the methodology

Thomas P. A. Debray,[a,b]* Karel G. M. Moons,[a,b] Gert van Valkenhoef,[c] Orestis Efthimiou,[d] Noemi Hummel,[e] Rolf H. H. Groenwold,[a] Johannes B. Reitsma[a,b] and on behalf of the GetReal methods review group

Individual participant data (IPD) meta-analysis is an increasingly used approach for synthesizing and investigating treatment effect estimates. Over the past few years, numerous methods for conducting an IPD meta-analysis (IPD-MA) have been proposed, often making different assumptions and modeling choices while addressing a similar research question. We conducted a literature review to provide an overview of methods for performing an IPD-MA using evidence from clinical trials or non-randomized studies when investigating treatment efficacy. With this review, we aim to assist researchers in choosing the appropriate methods and provide recommendations on their implementation when planning and conducting an IPD-MA. © 2015 The Authors. *Research Synthesis Methods* published by John Wiley & Sons, Ltd.

Keywords: meta-analysis; IPD; evidence synthesis; review; RCT; non-randomized intervention studies; NRSI; cross-design

1. Introduction

The evaluation of a novel drug or intervention typically involves a series of randomized clinical trials (RCTs) where its safety and efficacy are extensively tested. Because trials are often relatively small and typically exhibit differences in study design, selection of subjects, studied outcome(s), dosage, choice of comparator intervention, and quality of the conducted research, conflicting evidence occasionally arises. As a consequence, systematic reviews have become an important tool to summarize the evidence from these trials and to generalize their conclusions beyond their specific settings.

Over the past few decades, several methods have been developed to quantify the results of a systematic review. Most of these methods adopt a meta-analytical rationale and pool the results from individual studies by accounting for various forms of uncertainty (Sutton *et al*., 2009; Sutton and Higgins, 2008). Hereby, group-level summary statistics (aggregate data (AD)) that quantify the treatments' relative efficacy or safety are retrieved from the published literature or from study authors and are subsequently synthesized into a weighted average.

[a]*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands*
[b]*The Dutch Cochrane Centre, Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands*
[c]*Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*
[d]*Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina, Greece*
[e]*Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland*
**Correspondence to: Thomas P. A. Debray, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands.*
E-mail: T.Debray@umcutrecht.nl

An overview of the members of the GetReal methods review group is given in Supporting Information 1.

Unfortunately, when synthesizing published AD, even rigorously conducted meta-analyses can be of limited value. In particular, when there is substantial heterogeneity in estimates of relative treatment effect, a weighted average may no longer be informative in medical care. In such situations, it is important to identify whether treatment effects vary across clinical subgroups because of effect modification. Although published AD can be used for exploring modifiers of treatment effect, such approach lacks power when published summary statistics (e.g., mean age) do not vary much across studies. More importantly, the use of published AD to investigate effect modification is prone to bias because it cannot properly take subject-level characteristics into account. This bias is also known as the "ecological fallacy" (Berlin *et al*., 2002). Additional problems arise when AD are not available, poorly reported, derived and presented differently across studies (for example, odds ratio versus relative risk), and more likely to be reported (and in greater detail) when statistically or clinically significant (Riley *et al*., 2010). For this reason, investigators increasingly embark into an individual participant data meta-analysis (IPD-MA) (Riley *et al*., 2010; Thompson, 2009; Stewart and Tierney, 2002). These meta-analyses include the raw data from each relevant study that is (ideally) identified through a systematic review. By securing IPD of individual trials, it becomes possible to disentangle subject-level and study-level sources of heterogeneity in treatment effect (van Walraven, 2010; Michiels *et al*., 2005; Lyman and Kuderer, 2005). This, in turn, may help to explore effect modification (Simmonds and Higgins, 2007; Stewart and Tierney, 2002) and to consistently adjust for confounding variables (i.e., for differences in baseline characteristics across different treatment arms). Access to IPD may also help to improve data quality (Lyman and Kuderer, 2005; Stewart and Tierney, 2002; Lau *et al*., 1998), to standardize definitions and analyses (van Walraven, 2010; Lyman and Kuderer, 2005), to obtain complete follow-up data on all randomized participants (Clarke and Stewart, 1994), to combine studies with different follow-up times (Sud and Douketis, 2009; Lyman and Kuderer, 2005; Stewart and Tierney, 2002; Lau *et al*., 1998), to analyze multiple outcomes (Bujkiewicz *et al*., 2014; Li and Meredith, 2003), to investigate long-term outcomes (Stewart and Tierney, 2002), and to investigate rare exposures.

Because limited guidance currently exists on dealing with challenges that are specific to IPD-MA, we embarked on a literature review to identify state-of-the-art methods and to provide recommendations on their implementation. This review aims to help systematic reviewers with a limited background in medical statistics in identifying relevant methods for their IPD-MA.

2. Methods

2.1. Information sources

We conducted a literature review to identify scientific, peer-reviewed articles that focus on methods for IPD-MA to investigate treatment efficacy using evidence from clinical trials or non-randomized studies. Medline and Embase databases were queried using a combination of relevant keywords in the Pubmed and Ovid platform (Supporting Information 2). The corresponding search strategy was independently reviewed by two librarians. They concluded that the search strategy did not meaningfully improve when medical subject headings terms were included or search terms were truncated. A selected set of key journals was also hand-searched: the Journal of Research Synthesis Methods and the Journal of the Royal Statistical Society (series A, B, and C). Databases were searched from the earliest date until January 27, 2014. Finally, a last attempt to identify missed key articles was made by contacting key researchers in the field and performing cross-reference checks. Here, relevant hits were added until March 1, 2015.

Because the search for IPD methods was limited to medical and natural science databases, this review is not intended to be comprehensive. It is, for instance, well known that several IPD-MA methods have been developed in the field of social sciences. In general, the concepts of meta-analysis are universally applicable across different scientific domains, whereas the nature of the data may require tailoring of statistical methods. As a consequence, a more extensive search strategy is unlikely to add relevant variations in the identified methods and may fail to yield any new insights. This effect is known as theoretical saturation (Lilford *et al*., 2001).

2.2. Eligibility criteria

Articles were considered eligible for inclusion if they were written in English and addressed issues related to IPD-MA in intervention research. Key IPD-MA issues were defined as statistical models, software routines, simulation studies, empirical comparisons, didactic examples, and best practice guidelines. Articles were excluded when IPD-MA methods were applied to address solely a substantive clinical question, without a methodological focus. Articles for which no full text was available were also excluded.

2.3. Search results

A flowchart of the articles identified from our search is presented in Figure 1. A total of 3360 unique records were identified using the search strategy and were deemed eligible for title screening. Out of these, 3206 records were excluded because they did not have a methodological focus or did not correspond to a peer-reviewed

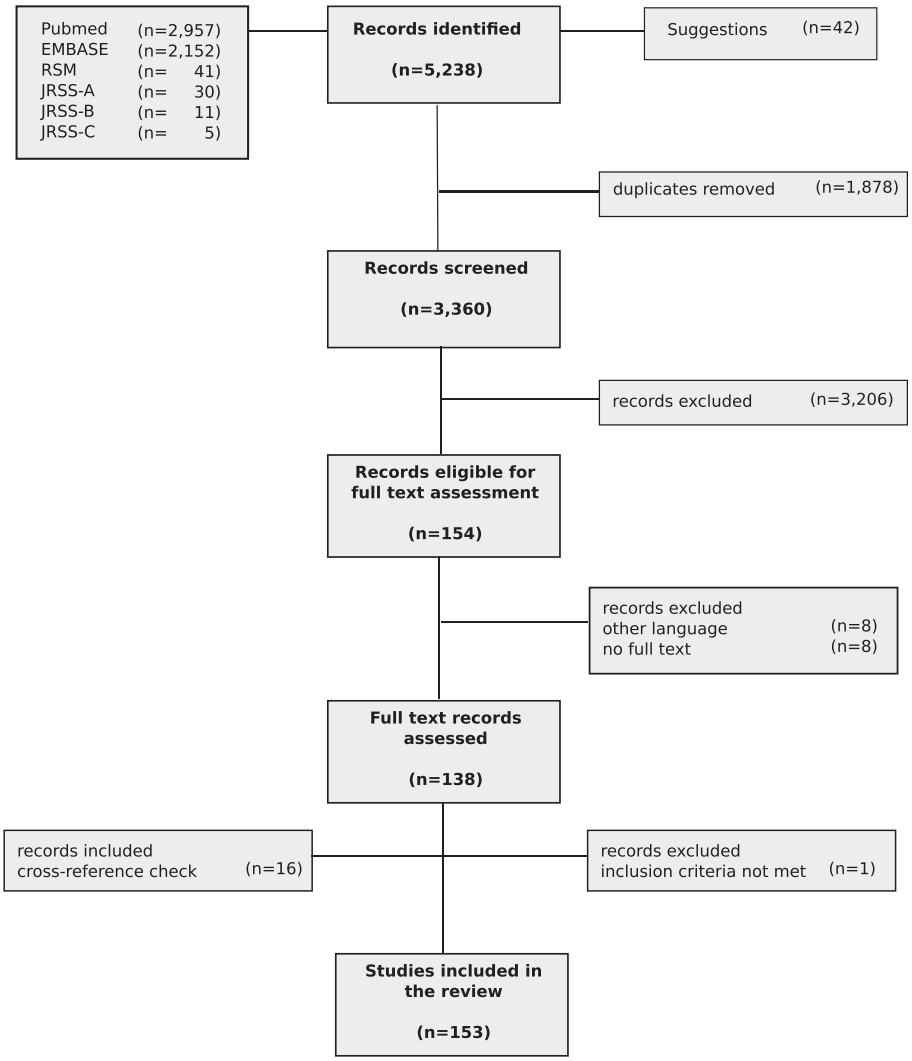

Figure 1. Flow diagram of selected studies.

article (e.g., conference poster). From the remaining 154 articles, another 16 records were excluded because of missing full text or non-English language. Finally, when assessing full text, 16 records were added after cross-reference checks, and one article was removed because it did not adhere to inclusion criteria. A total of 153 articles were included in the review (Supporting Information 4); a complete list of the included articles can be accessed online at https://www.zotero.org/groups/wp4_-_ipd_meta-analysis.

3. Results

3.1. Conceptual issues

The IPD from a single RCT are often analyzed using regression models where the observed outcome is modeled as a function of the subjects' undergone treatment (and, in many cases, other covariates of interest). The intercept term of these models then represents the study effect (e.g., baseline risk), whereas their regression coefficient represents the treatment effect. Regression models have two major advantages: First, their implementation is possible for numerous outcome types (e.g., continuous, binary, and time-to-event), and second, their estimated coefficients can be interpreted in a fairly straightforward manner. Unfortunately, the implementation of traditional regression models becomes challenging when a trial consists of multiple centers or when multiple trials need to be analyzed (as in an IPD-MA). In those situations, regression models need to account for clustering of subjects within trials (or within centers) by allowing each trial (or center) to have their "own" study and treatment effect. This approach becomes infeasible when individual trials (or centers) are relatively small. Moreover, corresponding regression models no longer yield an overall estimate of relative treatment effect, casting doubt about whether and under what circumstances a certain intervention is advantageous.

Meta-analysis attempts to overcome these issues by pooling relative treatment effects from multiple trials or centers (henceforth *studies*) and by borrowing information across these studies during the process. Over the past few decades, two alternate approaches have been proposed for conducting an IPD-MA (Stewart *et al*., 2012; Riley *et al*., 2010). The most common and conceptually least complicated approach is the so-called *two-stage approach*. In this approach, the IPD are first analyzed separately in each study to produce study-specific estimates of relative treatment effect (and possible treatment-covariate interactions). A combined estimate is then obtained in the second step by calculating a weighted average (e.g., inverse error-variance-based) of the individual estimates. Hereby, one may assume that all studies share a common treatment effect and that differences between estimates from the included studies solely arise because of sampling variation (fixed effects meta-analysis). Alternatively, it is possible to allow for heterogeneous treatment effects by incorporating variation between studies and performing a random effects meta-analysis.

In the so-called *one-stage approach*, the IPD from all studies are analyzed simultaneously by adopting a single statistical model that fully accounts (random effects meta-analysis), partially accounts (mixed effects meta-analysis), or does not account (fixed effects meta-analysis) for heterogeneity across studies. These single-stage models are more commonly known as multilevel or hierarchical models (Simmonds *et al*., 2005). The implementation of some common one-stage and two-stage meta-analysis model is illustrated in Supporting Information 3.

It is conventionally believed that the one-stage and two-stage approaches yield similar estimates of treatment effects (Stewart *et al*., 2012; Koopman *et al*., 2008b; Tudur Smith and Williamson, 2007; Steinberg *et al*., 1997). Indeed, it has been shown that under certain conditions, the one-stage and two-stage approaches may lead to equivalent results (Lyman and Kuderer, 2005; Mathew and Nordström, 2010), particularly when interest lies in estimating a single treatment effect estimate. The two-stage approach offers several advantages favoring its implementation in the medical literature. In particular, the two-stage approach is conceptually more intuitive and therefore requires less statistical expertise when conducting or interpreting an IPD-MA (Stewart *et al*., 2012; Koopman *et al*., 2007). Furthermore, because the two-stage approach analyzes each study separately, study-specific estimates are not influenced by external information. This discrepancy may, for instance, become relevant when effect modification varies across studies and corresponding associations do not follow a well-defined distribution (Piepho *et al*., 2012).

Unfortunately, the two-stage approach is known to have little power for detecting nonlinear associations between continuous exposures and the outcome(s) of interest and for detecting treatment-covariate interactions (Simmonds and Higgins, 2007; Simmonds *et al*., 2005; Tudur Smith *et al*., 2005b; Schmid *et al*., 2004). In addition, the two-stage approach may lead to bias in pooled effects, standard errors, between-study heterogeneity, and correlation between random effects when few studies or few participants (or events) per study are available (Debray *et al*., 2013; Stijnen *et al*., 2010; Tobías *et al*., 2004), when statistical models cannot fully account for follow-up times (Poppe *et al*., 2011; Lyman and Kuderer, 2005; Duchateau *et al*., 2001; Buyse and Piedbois, 1996) or for the time between recurrent events (Haines and Hill, 2011; Jones *et al*., 2009). Although methodology to overcome some of these limitations has been described (Stijnen *et al*., 2010), the one-stage approach offers the highest degree of flexibility for making necessary assumptions and is therefore often considered to be superior. The two-stage approach can, however, still be useful to explore the available data, to present intermediate results or to identify key challenges when designing a one-stage meta-analytical model (Stewart *et al*., 2012).

Although IPD-MA is often considered as gold standard, their implementation is no panacea against the limitations of meta-analyses that are solely based on published AD. In particular, meta-analysis of raw data may still be difficult when randomization or follow-up procedures vary, when studies have measured different covariates or when relevant descriptions of samples, settings, and treatments are not available for every study included in the meta-analysis (Burdett and Stewart, 2002). Furthermore, IPD-MA may become prone to bias when IPD are only sought for a specific subset of studies (e.g., studies conducted by contacts or friends in the research field) or when the availability of IPD is related to study results (Ahmed *et al*., 2012).

Several recommendations have been made to accommodate for these potential sources of bias (Ahmed *et al*., 2012; Stewart and Tierney, 2002). First of all, meta-analyses should be based on studies that are identified through a systematic review. Secondly, clinicians from the field should actively be involved as they may help to identify whether important studies may have been missed (Clarke and Godwin, 1998), to collect additional follow-up data where relevant and to resolve inconsistencies between different data sources. Finally, researchers should seek IPD for all relevant studies identified and collect AD for studies where IPD could not be obtained. Methods for combining IPD and AD are discussed in Section 3.5.

3.2. Statistical models to estimate an overall summary of treatment effect

In the succeeding texts, we describe the common methodology of *one-stage* IPD-MA models. These hierarchical models can be viewed as a direct extension of traditional regression models and are also known as generalized linear mixed models (GLMMs). In contrast to traditional regression models, hierarchical regression models allow coefficients to follow a certain distribution. For instance, GLMMs may specify that the relative treatment effect between two specific interventions varies across studies according to a normal distribution. The mean of this

distribution then simply represents the "average" treatment effect between the two interventions, and its variance indicates the degree of between-study heterogeneity in treatment effect.

Consider an IPD-MA of $i = 1, \ldots, M$ independent studies with $k = 1, \ldots, N_i$ subjects each. Let x_{ik} be a dummy variable that indicates treatment group (treatment or control) of subject k in study i. We denote the corresponding outcome as y_{ik}. This outcome can be continuous, binary, ordinal, count, and so on. The GLMM can then be stated as follows:

$$g(E(y_{ik})) = \alpha_i + \beta_i x_{ik} \tag{1}$$

In this model, $E(y_{ik})$ denotes the expected value of y_{ik}, the parameter α_i represents the study effect (e.g., baseline risk), and the parameter β_i represents the treatment effect. The link function g may take different forms depending on the type of outcome data. A detailed overview of possible implementations is discussed in Section 3.4 and also illustrated in Table 1.

When specifying the study effects in model (1), each α_i may be taken as fixed effects (estimated separately in each study), as a common effect (so $\alpha_i = \alpha$ for all studies) or as random effects (α_i is drawn from a certain distribution). Researchers typically allow for heterogeneous study effects by estimating a separate intercept α_i in each study (Abo-Zaid *et al.*, 2013; Higgins *et al.*, 2001; Thompson *et al.*, 2001; Turner *et al.*, 2000) or by specifying a random-effects distribution on α (Broström and Holmberg, 2011; Simmonds and Higgins, 2007; Goldstein *et al.*, 2002; Thompson *et al.*, 2001; Turner *et al.*, 2000). A random-effects distribution may also be specified for β_i when heterogeneity in treatment effects across studies is plausible (Abo-Zaid *et al.*, 2013; Bowden *et al.*, 2011; Katsahian *et al.*, 2008; Goldstein *et al.*, 2002; Higgins *et al.*, 2001; Thompson *et al.*, 2001). It is conventional and computationally efficient to assume normal distributions for the random effects (Sutton and Higgins, 2008; Thompson *et al.*, 2001). Other distributions are also possible (Broström and Holmberg, 2011; Thompson *et al.*, 2001) but often require more computational power and occasionally lead to convergence issues.

Table 1. Basic statistical models for estimating overall treatment effect

Outcome type	Model type	Basic statistical model	–
Continuous	GLMM	$y_{ik} \sim N(\mu_{ik}, \sigma_i^2)$ or $N(\mu_{ik}, \sigma^2)$ $\mu_{ik} = \alpha_i + \beta_i x_{ik}$	(L1)
Binary	GLMM	$y_{ik} \sim \text{Bernoulli}(p_{ik})$ $\text{logit}(p_{ik}) = \alpha_i + \beta_i x_{ik}$	(L2)
Ordinal	GLMM	$y_{ijk} \sim \text{Bernoulli}(q_{ijk})$ $\text{logit}(q_{ijk}) = \alpha_i + \zeta_j + \beta_i x_{ik}$	(L3a)
	GLMM	$y_{ijk} \sim \text{Bernoulli}(q_{ijk})$ $\text{logit}(q_{ijk}) = \alpha_{ij} + \beta_i x_{ik}$	(L3b)
Count	GLMM	$y_{ik} \sim \text{Poisson}(\mu_{ik})$ $ln(\mu_{ik}) = \alpha_i + \beta_i x_{ik}$	(L4)
Time-to-event	Cox PH	$h_{ik}(t) = h_{i0}(t)\exp(\beta_i x_{ik})$	(L5a)
	Cox PH	$h_{ik}(t) = h_0(t)\exp(\alpha_i + \beta_i x_{ik})$ where $\alpha_i \sim N(0, \sigma^2)$ or $\exp(\alpha_i) \sim \sigma^2\Gamma(\sigma^{-2})$	(L5b)
	Cox PH	$h_{ik}(t) = h_{i0}(t)\exp(\alpha_i + \beta_i x_{ik})$ with $\sum\alpha_i = 0$	(L5c)
	GLMM	$y_{ijk} \sim \text{Poisson}\left(\mu_{ijk}\right)$ $ln\left(\mu_{ijk}\right) = \alpha_i + \beta_i x_{ik} + \lambda_j + ln\left(t_{ijk}\right)$	(L5d)

GLMM, generalized linear mixed model; PH, proportional hazards.
Overview of statistical models for different outcome types. Each model is discussed in Section 3.4.

For instance, consider an IPD-MA with a binary outcome where study effects and treatment effects differ across studies. The following GLMM could then be specified to estimate an overall treatment effect that is adjusted for the presence of between-study heterogeneity:

$$\begin{gathered} y_{ik} \sim \text{Bernoulli}(p_{ik}) \\ \text{logit}(p_{ik}) = \alpha_i + \beta_i x_{ik} \,. \\ \beta_i \sim N(\mu, \tau) \end{gathered}$$

In this model, the pooled treatment effect (i.e., the "average" treatment effect across studies) is given by μ, and the degree of between-study heterogeneity in treatment effect is given by τ. Because the study effects are estimated separately in each study, there is no pooled study effect.

3.3. Statistical models to investigate heterogeneity in treatment effect across and within studies

In the previous section, we described common methods to produce an overall summary of treatment effect. These methods make little or no attempt to explain or investigate differences between study results (heterogeneity) across studies or to identify subgroup effects (Fisher *et al*., 2011; Sutton and Higgins, 2008; Simmonds and Higgins, 2007). There has been increasing interest in using meta-analysis to go beyond estimating only the overall treatment effect. As a consequence, meta-analyses increasingly attempt to identify modifiers of treatment effect by exploring the presence of interaction effects.

Traditional approaches such as meta-regression and subgroup analysis investigate the presence of *trial-level* interaction, that is, interaction between treatment status and a specific study-level covariate. This covariate may represent a certain study characteristic (such as level of blinding) or a summarized subject-level characteristic (such as mean age) (Fisher *et al*., 2011; Sutton *et al*., 2008; Tudur Smith *et al*., 2005b; Li and Meredith, 2003; Yamaguchi *et al*., 2002). For instance, Berlin *et al.* (2002) used meta-regression to identify whether anti-lymphocyte antibody induction therapy is more beneficial in patients with elevated panel reactive antibodies (PRA). Hereto, they determined the percent of patients with elevated PRA within each study and estimated their association with the corresponding treatment's effect size (Berlin *et al*., 2002).

Although *trial-level* interactions may indicate the presence of effect modification, it has been demonstrated that such interactions have low statistical power for identifying modifiers of treatment effect and may lead to ecological (aggregation) bias. In particular, associations between aggregated values may not be representative for individual subjects (Lambert *et al*., 2002). In the example of induction, meta-regression failed to identify effect modification by PRA because the level of benefit within PRA category changed across studies. A remedy against this pitfall is to use IPD and to investigate the presence of *subject-level* interaction (rather than trial-level interaction). This can be achieved by specifying an interaction term between treatment status and subject-level covariate in model (1). Details on how to do this in one-stage and two-stage approaches are provided in Supporting Information 3. The inclusion of interaction terms may also help to adjust for baseline imbalances despite randomization (Higgins *et al*., 2001) and to increase the precision of treatment effect estimates (by including a strong prognostic factor). Estimated interaction terms must, however, be interpreted with care as they describe a mixture of trial-level and subject-level interactions (Higgins *et al*., 2001). For this reason, when a modifier of interest has sufficient variation within and across studies, researchers sometimes specify two (instead of one) interaction terms: one interaction between treatment status and the study-specific mean of the covariate (i.e., the values that would be used in meta-regression) and one interaction between treatment status and the study-centered covariate values (Donegan *et al*., 2012; Fisher *et al*., 2011). In this manner, it becomes possible to quantify the presence of ecological bias.

Individual participant data meta-analysis often assume common effects for interaction terms to restrict model complexity and facilitate estimation (Crowther *et al*., 2012; Teramukai *et al*., 2004). It is, however, possible that interaction terms are actually heterogeneous across studies. For this reason, some researchers propose to estimate subject-level interactions in isolation ("meta-analysis of interaction estimates"). This can be achieved by adopting a two-stage approach where interaction effects are estimated separately within each study and then pooled in the second stage of the IPD-MA (Fisher *et al*., 2011; Simmonds and Higgins, 2007). Alternatively, it is possible to adopt a one-stage approach by assuming random-effects distributions for the interaction effects (Simmonds and Higgins, 2007; Schmid *et al*., 2004). The latter strategy has been demonstrated to have the greatest power and flexibility, even when few studies are available, and is therefore generally recommended (Katsahian *et al*., 2008; Simmonds and Higgins, 2007). Finally, researchers may also attempt to model nonlinear interactions, for example, by using fractional polynomials (Royston and Sauerbrei, 2004), to model two-way interactions between treatment and study-level covariates or even to model three-way interactions between treatment, subject-level covariates, and study-level covariates.

It is important to realize that practical implementation of IPD-MA with interaction terms requires careful thought, as the range of possible analyses (i.e., options for treating the unknown parameters) could easily lead to over-fitting and data-dredging. This problem becomes even more apparent when multiple study-level and

subject-level covariates are available and IPD are limited. For this reason, it is recommended to pre-specify statistical models in a study protocol before collecting or analyzing the IPD (Higgins *et al*., 2001).

3.4. *Modeling of specific types of outcomes*

In the succeeding texts, we describe common statistical models for analyzing specific types of outcomes and describe typical estimation procedures. An overview is provided in Table 1.

3.4.1. Continuous. Continuous outcome data are typically modeled using an identity link function (L1). In this model, β_i indicates the absolute change in outcome due to treatment for patients from the i^{th} study. If the outcome of interest represents a change from baseline (e.g., a change in depression score since inclusion by the trial), it is appropriate to adjust for baseline responses. Further discussion can be found in Supporting Information 3 and (Riley *et al*., 2013; Goldstein *et al*., 2002; Higgins *et al*., 2001; Thompson *et al*., 2001; Goldstein *et al*., 2000).

3.4.2. Binary. When modeling a binary outcome, it is common to use a logit link function (L2). In this model, $\exp(\beta_i)$ represents the odds ratio for the treatment effect in the i^{th} study. If β_i is assumed as a common effect ($\beta_i = \beta$ for all studies) or as a random effect (e.g., $\beta_i \sim N(\beta, \tau^2)$), then $\exp(\beta)$ represents the odds ratio for the overall treatment effect. Examples are provided by (Thomas *et al*., 2014; Debray *et al*., 2013; Stijnen *et al*., 2010; Sutton *et al*., 2008; Thompson *et al*., 2001; Wakefield and Salway, 2001; Turner *et al*., 2000).

3.4.3. Ordinal. The logistic model can easily be extended to account for ordinal outcome-type data by relating to the proportional odds model (Thompson *et al*., 2001; Whitehead *et al*., 2001). Let q_{ijk} denote the probability of subject k in study i having a response in category j or below. Hereby, it is assumed that the categories are ordered in terms of desirability, thus, lower categories are better. The proportional odds model is described in (L3a), where the parameter ζ_j represents the log odds ratios for each category cutoff. This model assumes that there is a common treatment effect β_i and a common study effect α_i within each treatment group. It is possible to relax these assumptions by allowing study effects to vary across different treatment groups (L3b). When the assumption of proportional odds is violated, generalized ordered models or partial proportional models may be used instead of the proportional odds model.

3.4.4. Count. Denote y_{ik} as the number of events for subject i in study k. A straightforward one-stage approach may adopt a Poisson distribution with a log-link (L4). Several extensions have been described that for instance, allow for data with many zeros by introducing a discrete point mass (hierarchical zero-inflated Poisson regression) (Lee *et al*., 2006; Yau and Lee, 2001; Hall, 2000).

3.4.5. Time-to-event. Let t_{ik} denote the observed time (either censoring time or event time) for subject k in study i. The outcome y_{ik} then represents an indicator that the time corresponds to an event ($y_{ik} = 1$) or a censoring time ($y_{ik} = 0$). Common multilevel survival models distinguish between heterogeneity in baseline hazard (e.g., due to differences in incidence of the event between studies given the same treatment) and heterogeneity in relative treatment effect (e.g., due to differences in treatment effect size) (Michiels *et al*., 2005).

The hierarchical Cox proportional hazards regression model is a popular approach for analyzing survival data from multiple studies as it requires no assumptions regarding the distribution of the baseline hazard rate (Michiels *et al*., 2005; Tudur Smith *et al*., 2005a,2005b; Sargent, 1998). This model may account for heterogeneity in baseline hazard (that is, the presence of study effects) by stratifying the baseline hazard by study (L5a). Examples of this strategy are provided in (Crowther *et al*., 2012; Bowden *et al*., 2011; Thompson *et al*., 2010; Katsahian *et al*., 2008; Tudur Smith and Williamson, 2007; Michiels *et al*., 2005; Tudur Smith *et al*., 2005a,2005b; Trikalinos and Ioannidis, 2001). Because L5a does not yield a direct estimate of the trial effect, more advanced approaches attempt to investigate this heterogeneity by introducing a frailty term α_i. It is then assumed that the hazards within each study are proportional to the same common baseline hazard function (L5b). Examples of this second strategy are provided in (Crowther *et al*., 2012; Bowden *et al*., 2011; Katsahian *et al*., 2008; Rondeau *et al*., 2008; Tudur Smith *et al*., 2005a; Yamaguchi *et al*., 2002; Vaida and Xu, 2000; Yamaguchi and Ohashi, 1999; Sargent, 1998; Yau and McGilchrist, 1998). Alternatively, it may be assumed that the baseline hazard functions for each study have the same shape but may have different magnitudes (L5c) (Michiels *et al*., 2005). Further extensions of these models may consider time-dependent frailties (Yau and McGilchrist, 1998).

Similar to other outcome-type data, heterogeneity in treatment effect can be investigated by specifying a random effects distribution, for example, $\beta_i \sim N(\beta, \tau^2)$ or $\beta_i = \beta + b_i$ where $b_i \sim N(0, \tau^2)$ (Rondeau *et al*., 2008; Michiels *et al*., 2005; Tudur Smith *et al*., 2005a; Yamaguchi *et al*., 2002; Vaida and Xu, 2000; Yamaguchi and Ohashi, 1999). In general, the use of stratified models with random treatment effects (that is, L5a or L5c using normally distributed frailty terms) has been recommended as these models maintain the within-study structure while allowing for heterogeneity in study and treatment effects (Tudur Smith *et al*., 2005a; Yamaguchi *et al*., 2002).

Unfortunately, there are significant computational challenges in the fitting of random treatment effects to time-to-event outcome data (Crowther *et al*., 2012; Fisher *et al*., 2011; Thompson *et al*., 2010; Sargent, 1998). For this reason, researchers may choose to approximate the baseline hazard function by defining a piece-wise constant or a spline function (Royston and Parmar, 2002; Vaida and Xu, 2000; Sargent, 1998; Yau and McGilchrist, 1998). Alternatively, it has been shown that the Cox proportional hazards model can be approximated with a Poisson GLMM by splitting follow-up time t_{ik} (and event indicator y_{ik}) into j narrow time intervals of fixed length (L5d) (Crowther *et al*., 2012). The parameter λ_j then represents the baseline hazard rate during the j^{th} time interval, and β is once again the log hazard ratio for the treatment effect. An additional advantage of this approach is that the proportional hazards assumption can be relaxed, for example, by replacing the Poisson with a log-gamma distribution or by pre-specifying the baseline function $h_0(t)$ in L5b (Barrett *et al*., 2012; Siannis *et al*., 2010).

3.4.6. Other. Several one-stage models have been proposed for IPD-MA where the outcome belongs to another type of non-normally distributed data (Thompson *et al*., 2001), recurrent events (Haines and Hill, 2011), repeated measurements (Jones *et al*., 2009), or multivariate responses (Kim *et al*., 2013; Riley *et al*., 2008; Yamaguchi *et al*., 2002).

3.5. Meta-analysis of individual participant data combined with aggregate data

Subject-level data are often unavailable for all relevant studies (Ravva *et al*., 2014; Donegan *et al*., 2013; Saramago *et al*., 2012; Jansen, 2012; Riley *et al*., 2008; Sutton *et al*., 2008; Riley *et al*., 2007). Therefore, published AD may represent an additional source of evidence. Including published AD in an IPD-MA is strongly recommended when studies providing IPD systematically differ from studies for which IPD were not obtained, as this may lead to data availability bias or reviewer selection bias (Ahmed *et al*., 2012). But even when there is no indication of such bias, inclusion of AD in an IPD-MA for studies of which IPD are lacking may be considered to increase the statistical power for detecting treatment effects or treatment-covariate interactions (Donegan *et al*., 2013; Jansen, 2012; Saramago *et al*., 2012). In general, there are three alternate approaches for combining IPD and AD in a meta-analysis:

- Meta-analysis of AD

In this approach, the available IPD are first reduced to AD (Ravva *et al*., 2014; Riley *et al*., 2007) and then pooled into a weighted average as in two-stage meta-analysis. When investigating the presence of effect modification, caution is warranted to avoid the introduction of ecological bias (Signorovitch *et al*., 2012; Jackson *et al*., 2008; Riley *et al*., 2008, 2007; Tudur Smith *et al*., 2005b; Wakefield and Salway, 2001).

- Meta-analysis of reconstructed IPD

It is possible to reconstruct IPD from published aggregate information for binary (based on 2×2 tables), ordinal (based on the number of responses within each treatment category), and survival type data (based on Kaplan–Meier survival curves) (Guyot *et al*., 2012; Riley and Steyerberg, 2010; Riley *et al*., 2007). In particular, this summary information can be used to restore subject-level information on treatment group and outcome status (or time until event). Afterwards, (traditional) one-stage models can be used for evidence synthesis. Unfortunately, reconstructed IPD do not have subject-level information on covariates and may therefore lead to ecological bias when investigating the presence of effect modification.

- Hierarchical-related regression

This approach directly combines the likelihood from each data source (IPD or AD), for example, using Bernoulli (IPD) and binomial (AD) distributions for modeling the individual responses and trial-specific summaries of binary outcome data (Ravva *et al*., 2014; Saramago *et al*., 2012; Jansen, 2012; Sutton *et al*., 2008; Jackson *et al*., 2008, 2006; Goldstein *et al*., 2002). In this manner, it ensures that all studies contribute toward estimation of the overall treatment effect and study-level covariates and that only IPD are used to estimate subject-level covariate effects. Hierarchical related regression is also known as shared parameters models (Donegan *et al*., 2013) and can, for instance, be achieved by fitting a multilevel model that includes a dummy variable indicating whether a data source relates to IPD or AD (Riley *et al*., 2008, 2007; Goldstein *et al*., 2000).

Unfortunately, the incorporation of AD in an IPD-MA does not necessarily improve the validity of estimated treatment effects and interactions. In particular, amongst other issues, the exclusion of subjects post randomization (Tierney and Stewart, 2005) as well as the presence of limited follow-up data and informative drop-out may introduce bias in published AD. Because researchers may often find themselves between the Scylla and Charybdis when deciding to include published AD in an IPD-MA, it has been recommended to assess the sensitivity of meta-analysis results to the inclusion of AD (Riley and Steyerberg, 2010; Riley *et al*., 2008, 2007). This can, for instance, be achieved by performing an IPD-only meta-analysis and meta-analysis that combines IPD and AD. An additional advantage of sensitivity analyses is that they may help readers without (advanced) statistical expertise to interpret results in function of the underlying assumptions.

3.6. Multiple treatment comparisons

In a different article of this series, we discuss in depth the use of network meta-analysis (NMA) models for synthesizing information from multiple treatments across different studies (Efthimiou *et al*., 2015). These models allow the estimation of the relative treatment effect between many (rather than two) specific interventions, even when these interventions have never been compared head-to-head. The implementation of NMA models is, however, less straightforward than traditional meta-analysis models and requires studies to have similar covariate distributions for important effect modifiers (Signorovitch *et al*., 2010). For this reason, it is important to correct for baseline imbalance between studies (Ali *et al*., 2013; Jansen, 2012) and explore the presence of treatment-covariate interactions (Donegan *et al*., 2012; Jansen, 2012; Signorovitch *et al*., 2012). This is ideally achieved using IPD.

The specification of IPD-NMA models can be viewed as an extension of regression model (1) where a dummy variable is included for each (except one) treatment (Higgins *et al*., 2001). It is, however, more common to describe NMA models as a function of consistency equations, which simplifies the specification of interaction effects (Donegan *et al*., 2013; Jansen, 2012; Saramago *et al*., 2012). The use of IPD in NMA has been advocated even if IPD are only available for a fraction of the studies (Jansen, 2012). Recently, several NMA models have been proposed that allow the combined use of IPD and AD in a joint analysis (Donegan *et al*., 2013; Jansen, 2012; Saramago *et al*., 2012). These models were shown to have improved network consistency when compared to NMA models that were solely based on AD. Similar to AD-NMA, the quality of evidence resulting from an IPD-NMA strongly depends upon the quality and choice of included studies, the complexity and credibility of statistical models, and the transparency of reporting. A questionnaire for assessing the relevance and credibility of NMA studies has recently been reported by the International Society for Pharmacoeconomics and Outcomes Research and may help to improve the quality of IPD-NMA studies (Jansen *et al*., 2014).

3.7. Cross-design synthesis

Most trials are conducted to establish efficacy and safety of a single treatment in a specific study setting, and therefore it is often difficult to translate this evidence to "real-world" effectiveness where different concerns may be at play (Prevost *et al*., 2000; Stuart *et al*., 2011). For this reason, there is a growing interest to overcome the weaknesses of individual study designs by carefully combining different types of evidence (Droitcour *et al*., 1993). In particular, intervention research might benefit by the inclusion of non-randomized study data such as from (case-)cohort studies, case-control studies, or patient registries (Higgins *et al*., 2013; Kaizar, 2011; Peters *et al*., 2005; Prevost *et al*., 2000; Li and Begg, 1994). This strategy is more commonly known as cross-design synthesis or generalized synthesis of evidence (Sutton and Higgins, 2008; Sutton *et al*., 1998).

Because non-randomized studies of interventions (NRSI) are inherently prone to confounding, there is substantial controversy about the magnitude and interpretation of potential differences between the results from RCTs and NRSI (Reeves *et al*., 2003; Valentine and Thompson, 2013; Schmidt *et al*., 2013; Golder *et al*., 2011; Ioannidis *et al*., 2001; Concato *et al*., 2000). Whether or not it is valid to combine RCT and NRSI in IPD-MA and under what conditions is an ongoing debate in the literature (Higgins *et al*., 2013; Reeves *et al*., 2003). So far, cross-design synthesis has only been advocated when the likelihood of conducting randomized trials is low (e.g., due to ethical concerns or practical limitations) or when research questions relate to unintended, rare (often harms), or long-term outcomes (benefits or harms). The justification of including NRSI may further increase when the derived estimates of relative treatment effects are deemed to be resistant against bias. This becomes more plausible when there are no systematic differences between the groups being compared (e.g., in terms of care provided, exposure to other factors, baseline characteristics, or in terms of how outcomes are determined).

To improve the validity of meta-analyses incorporating evidence resulting from NRSI, several suggestions have been made. First, researchers should critically appraise the risk of bias in available studies prior to cross-design synthesis and carefully decide whether inclusion is appropriate (Berger *et al*., 2014; Higgins *et al*., 2013; Reeves *et al*., 2003; Wells *et al*., 2013). Subsequently, researchers should routinely account for confounding factors within studies. Finally, researchers should explore potential sources of between-study heterogeneity and compare the results between randomized studies and NRSI (Valentine and Thompson, 2013).

3.8. Dealing with missing data

Missing data are a common problem when analyzing IPD from any (single) study and are typically classified into missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Rubin, 1987). These types correspond to situations where the probability of data being missing is completely independent of observed and unobserved factors (MCAR), depends only on the observed data (MAR), or also depends on unobserved data (MNAR) (Donders *et al*., 2006; Schafer, 1999; Rubin, 1987). When data are missing within a single study, it is generally recommended to apply multiple imputation strategies by adopting a statistical model that assumes MCAR or MNAR. We can distinguish the following missing data scenarios in an IPD-MA (Sutton *et al*., 1998):

- missing data for some subjects in one or more studies,
- missing data for all subjects of one or more studies (e.g., variables that have not been measured, outcomes that are missing, or missing study-level covariates), and
- entire study missing (e.g., when study authors are unable to share IPD).

Because the association between observed and missing variables may differ across studies, it has been recommended to allow for between-study heterogeneity during imputation. This can, for instance, be achieved by imputing IPD studies separately (two-stage imputation) (Burgess *et al*., 2013; Koopman *et al*., 2008a). Two-stage imputation is obviously not feasible if data are missing for all subjects within one or more studies. In such scenarios, it is desirable to use the so-called one-stage imputation models that borrow information across the available studies (Jolani *et al*., 2015; Burgess *et al*., 2013; Resche-Rigon *et al*., 2013; White *et al*., 2008; Yucel, 2008, 2011).

3.9. Software and estimation techniques for individual participant data meta-analysis

Several software packages exist for conducting an IPD-MA and adopt a frequentist or a Bayesian estimation framework (Tables 2 and 3). Frequentist methods typically maximize the likelihood function of the meta-analysis model using expectation–maximization (Vaida and Xu, 2000), Newton–Raphson, or Fisher scoring algorithms. Unfortunately, maximization of the (log-)likelihood function becomes problematic when there are few studies in the meta-analysis or included studies are small (Crowther *et al*., 2012; Li *et al*., 2011; Thompson *et al*., 2001; Higgins *et al*., 2001). For this reason, it is often recommended to adopt more advanced estimation techniques such as restricted (or residual) maximum likelihood estimation (Bowden *et al*., 2011; Higgins *et al*., 2001), penalized maximum likelihood estimation (Rondeau *et al*., 2008), penalized quasi-likelihood (Thompson *et al*., 2001), extended quasi-likelihood estimation, penalized partial likelihood estimation, or semi-parametric-penalized

Table 2. Software packages for fitting meta-analysis models

Software	Package	Characteristics	Used in
WinBUGS, JAGS, Stan, OpenBUGS	–	Fitting of one-stage and two-stage meta-analysis models using Bayesian Markov chain Monte Carlo (MCMC)	[26, 32, 48, 52, 67, 81, 86, 95, 98, 118, 126, 146, 147]
NONNEM	–	Fitting of GLMM	[87]
R, S-plus*	–	Unspecified	[9, 103, 104]
–	ecoreg	Estimation of individual-level covariate-outcome associations using AD ("ecological inference") or a combination of AD and individual participant data (IPD) ("hierarchical-related regression")	[52]
–	glmmML	Fitting of GLMM using ML. Allows for non-normal distributions in the specification of random intercepts.	[11]
–	hglm	Fitting of GLMM where the random effect may come from a conjugate exponential-family distribution	–
–	lme4	Fitting of GLMM using ML or REML (for mixed linear models only)	[11, 30, 56, 65, 110, 153], SI3
–	MASS	Fitting of GLMM using penalized quasi-likelihood (PQL)	–
–	mvmeta	Meta-analysis and meta-regression of AD (two-stage meta-analysis)	[30], SI3
–	nlme	Fitting of linear mixed-effects models using ML or REML	[89, 122]
–	survival	Fitting of Cox PH and mixed effect survival models using penalized partial likelihood estimation (PPL)	[44, 122, 132, 133]
–	frailtypack	Fitting of frailty models using semi-parametric-penalized likelihood (SPL)	[97]
–	coxme	Fitting mixed effects Cox PH models	–

GLMM, generalized linear mixed model; AD, aggregate data; PH, proportional hazards; REML, restricted (or residual) maximum likelihood; ML, maximum likelihood.
References are provided in Supporting Information (SI) 4.
*An overview of software packages for two-stage meta-analysis can be found on http://cran.r-project.org/web/views/MetaAnalysis.html.

Table 3. Software packages for fitting meta-analysis models

Software	Package	Characteristics	Used in
SAS	–	Unspecified	[106, 122, 132, 133]
–	PROC GLIMMIX	Fitting of GLMM using penalized quasi-likelihood (PQL)	[82, 129]
–	PROC GLM	Fitting of GLMM using MOM	[148]
–	PROC LOGISTIC	Fitting of mixed nonlinear models (binary/ordinal/nominal responses) using ML	[147]
–	PROC MIXED	Fitting of mixed linear models using ML, REML, or MOM. Can also perform two-stage meta-analysis.	[48, 57, 93]
–	PROC NLMIXED	Fitting of mixed nonlinear models using (approximated) ML	[95, 114, 148]
Stata	–	Unspecified	[3, 4, 45]
–	gllamm	Fitting of GLMM using ML	–
–	mvmeta	Meta-analysis and meta-regression of aggregate data (AD) (two-stage meta-analysis)	[5]
–	REGOPROB2	Fitting of random effects generalized-ordered probit models	[26, 95, 122]
–	stmixed	Fitting of flexible parametric survival models with mixed effects	–
–	XT	Fitting of GLMM	–
MLwiN, MLn	–	Fitting of GLMM and survival models using ML, REML, and EM	[41, 48, 122, 126, 136, 147]
FORTRAN, C	–	Unspecified	[60, 97, 137]

GLMM, generalized linear mixed model; REML, restricted (or residual) maximum likelihood; ML, maximum likelihood; MOM, method of moments; EM, expectation maximization.
References are provided in Supporting Information 4.

likelihood estimation. Unfortunately, these techniques are not widely implemented to investigate nonlinear outcomes (e.g., binary, ordinal, or count data).

The Bayesian framework combines the likelihood function with prior information (Yamaguchi *et al.*, 2002; Higgins *et al.*, 2001; Sargent, 1998; Larose and Dey, 1997) and is typically implemented within WinBUGS, JAGS, OpenBUGS, or Stan. The use of prior information may help to overcome convergence issues when studies are small or few studies are available (Table 4). Semi-informative prior distributions may, for instance, be used when model parameters are unidentifiable due to a lack of data (Jackson *et al.*, 2008). Alternatively, informative prior distributions may be used, for example, when relevant information can be borrowed from historical data

Table 4. Overview of prior distributions used within the Bayesian framework

–	Prior distribution	95% prior belief	Used in
Parameters for fixed effects (regression coefficients)			
Non-informative	$\beta \sim N(0, 10^6)$	$[0.00; +\infty)$ for the odds ratio	[98, 118]
–	$\beta \sim N(0, 10^5)$	$[0.00; +\infty)$ for the odds ratio	[26]
–	$\beta \sim N(0, 10^4)$	$[0.00; +\infty)$ for the odds ratio	[32, 48, 70, 86, 95, 126, 147]
–	$\beta \sim N(0, 10)$	[0.00; 494] for the odds ratio	[86]
Weakly informative	$\beta \sim N(0, 1.47)$	[0.20; 5.03] for the odds ratio	[52]
Parameters for between-study standard deviation			
Non-informative	$\tau^2 \sim \Gamma^{-1}(10^{-3}, 10^3)$	$[1.2 \times 10^7, +\infty)$ for τ	[67]
–	$\tau^2 \sim \Gamma^{-1}(10^{-3}, 10^{-3})$	$[3.2 \times 10^4; +\infty)$ for τ	[48, 81, 147]
–	$\tau^2 \sim U(0.01, 100)$	[1.62; 9.87] for τ	[70]
–	$\tau \sim N(0, 10)I(0)$	[0.30; 22.48] for τ	[146]
–	$\tau \sim U(0, 10)$	[0.25; 9.76] for τ	[32, 118]
Weakly informative	$\tau \sim U(0, 5)$	[0.12; 4.88] for τ	[86]
–	$\tau \sim U(0, 2)$	[0.05; 1.96] for τ	[98]
–	$\tau \sim N(0, 1)I(0)$	[0.03; 2.24] for τ	[26, 95]
–	$\tau \sim U(0, 1)$	[0.02; 0.98] for τ	[126]
–	$\tau \sim N(0, 0.125)I(0)$	[0.01; 0.80] for τ	[86]
–	$\tau^2 \sim \Gamma^{-1}(1, 10^{-2})$	[0.05; 0.62] for τ	[52]
–	$\tau \sim N(0, 0.033)I(0)$	[0.01; 0.41] for τ	[86]

References are provided in Supporting Information 4.

(Viele *et al.*, 2014). An additional advantage of the Bayesian framework is that it allows more flexibility in choosing and combining likelihood functions that are appropriate for the data that are being analyzed. As such, the use of Bayesian meta-analyses may sometimes help to improve the credibility of estimated treatment effects and between-study heterogeneity. The Bayesian framework, however, requires careful specification of prior distributions, investigation of convergence, and the drawing of inferences from a large simulated sample. Because it tends to be cumbersome and requires substantial statistical expertise, Bayesian IPD-MA is not pursued by many researchers.

4. Concluding remarks

Individual participant data meta-analyses offer numerous advantages over meta-analyses that are solely based on published AD and are therefore considered as a gold standard in evidence synthesis. Nevertheless, the presence of publication bias, data accessibility bias, or reviewer selection bias may still hamper or invalidate the results of an IPD-MA (Ahmed *et al.*, 2012), and synchronizing between different sources of IPD is not always straightforward. For this reason, it is not recommended to conduct IPD-MA on an *ad hoc* basis (e.g., without systematic review). Because the implementation of an IPD-MA requires additional efforts and statistical expertise, researchers should carefully assess whether the potential advantages (e.g., increased power, reduced bias, and investigating interaction and subgroup effects) outweigh the extra efforts involved. Finally, researchers should be aware that IPD-MA may have similar issues to meta-analyses of published AD and are no panacea against poorly designed and conducted primary research.

Acknowledgements

The research leading to these results was conducted as part of the GetReal consortium. For further information, please refer to www.imigetreal.eu.

Funding

The work leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement number 115546, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/20072013) and EFPIA companies' in-kind contribution

References

Abo-Zaid GMA, Guo B, Deeks JJ, Debray TPA, Steyerberg EW, Moons KGM, Riley RD 2013. Individual participant data meta-analyses should not ignore clustering. *Journal of Clinical Epidemiology* **66**: 865–873. DOI:10.1016/j.jclinepi.2012.12.017.

Ahmed I, Sutton AJ, Riley RD 2012. Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: a database survey. *British Medical Journal* **344**: d7762. DOI:10.1136/bmj.d7762.

Ali S, Mealing S, Hawkins N, Lescrauwaet B, Bjork S, Mantovani L, Lampertico P 2013. The use of individual patient-level data (IPD) to quantify the impact of pretreatment predictors of response to treatment in chronic hepatitis b patients. *BMJ Open* **3**: e001309. DOI:10.1136/bmjopen-2012-001309.

Barrett JK, Farewell VT, Siannis F, Tierney J, Higgins JPT 2012. Two-stage meta-analysis of survival data from individual participants using percentile ratios. *Statistics in Medicine* **31**: 4296–4308. DOI:10.1002/sim.5516.

Berger ML, Martin BC, Husereau D, Worley K, Allen JD, Yang W, Quon NC, Mullins CD, Kahler KH, Crown W 2014. A questionnaire to assess the relevance and credibility of observational studies to inform health care decision making: an ISPOR-AMCP-NPC good practice task force report. *Value in Health* **17**: 143–156. DOI:10.1016/j.jval.2013.12.011.

Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-lymphocyte antibody induction therapy study group. 2002. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine* **21**: 371–387. DOI: 10.1002/sim.1023.

Bowden J, Tierney JF, Simmonds M, Copas AJ, Higgins JP 2011. Individual patient data meta-analysis of time-to-event outcomes: one-stage versus two-stage approaches for estimating the hazard ratio under a random effects model. *Research Synthesis Methods* **2**: 150–162. DOI:10.1002/jrsm.45.

Broström G, Holmberg H 2011. Generalized linear models with clustered data: fixed and random effects models. *Computational Statistics & Data Analysis* **55**: 3123–3134. DOI:10.1016/j.csda.2011.06.011.

Bujkiewicz S, Thompson JR, Sutton AJ, Cooper NJ, Harrison MJ, Symmons DPM, Abrams KR 2014. Use of Bayesian multivariate meta-analysis to estimate the HAQ for mapping onto the EQ-5D questionnaire in rheumatoid arthritis. *Value in Health* **17**: 109–115. DOI:10.1016/j.jval.2013.11.005.

Burdett S, Stewart LA 2002. A comparison of the results of checked versus unchecked individual patient data meta-analyses. *International Journal of Technology Assessment in Health Care* **18**: 619–624.

Burgess S, White IR, Resche-Rigon M, Wood AM 2013. Combining multiple imputation and meta-analysis with individual participant data. *Statistics in Medicine* **32**: 4499–4514. DOI:10.1002/sim.5844.

Buyse M, Piedbois P 1996. On the relationship between response to treatment and survival time. *Statistics in Medicine* **15**: 2797–2812. DOI:10.1002/(SICI)1097-0258(19961230)15:24<2797::AID-SIM290>3.0.CO;2-V.

Clarke M, Godwin J 1998. Systematic reviews using individual patient data: a map for the minefields? *Annals of oncology: official journal of the European Society for Medical Oncology/ESMO* **9**: 827–833. DOI:10.1023/A:1008468705492.

Clarke MJ, Stewart LA 1994. Obtaining data from randomised controlled trials: how much do we need for reliable and informative meta-analyses? *British Medical Journal* **309**: 1007–1010. DOI:10.1136/bmj.309.6960.1007.

Concato J, Shah N, Horwitz RI 2000. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine* **342**: 1887–1892. DOI:10.1056/NEJM200006223422507.

Crowther MJ, Riley RD, Staessen JA, Wang J, Gueyffier F, Lambert PC 2012. Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Medical Research Methodology* **12**: 34. DOI:10.1186/1471-2288-12-34.

Debray TPA, Moons KGM, Abo-Zaid GMA, Koffijberg H, Riley RD 2013. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLoS One* **8**: e60650. DOI:10.1371/journal.pone.0060650.

Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM 2006. Review: a gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology* **59**: 1087–1091. DOI:10.1016/j.jclinepi.2006.01.014.

Donegan S, Williamson P, D'Alessandro U, Garner P, Smith CT 2013. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: individual patient data may be beneficial if only for a subset of trials. *Statistics in Medicine* **32**: 914–030. DOI:10.1002/sim.5584.

Donegan S, Williamson P, D'Alessandro U, Tudur Smith C 2012. Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: individual patient-level covariates versus aggregate trial-level covariates. *Statistics in Medicine* **31**: 3840–3857. DOI:10.1002/sim.5470.

Droitcour J, Silberman G, Chelimsky E 1993. A new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. *International Journal of Technology Assessment in Health Care* **9**: 440–449. DOI:10.1017/S0266462300004694.

Duchateau L, Pignon JP, Bijnens L, Bertin S, Bourhis J, Sylvester R 2001. Individual patient-versus literature-based meta-analysis of survival data: time to event and event rate at a particular time can make a di_erence, an example based on head and neck cancer. *Controlled Clinical Trials* **22**: 538–547. DOI:10.1016/S0197-2456(01)00152-0.

Efthimiou O, Debray TPA, van Valkenhoef G, Trelle S, Panayidou K, Moons KGM, Reitsma JB, Shang A, Salanti G, on behalf of the GetReal methods review group. 2015. Methods for network meta-analysis: a review. *Research Synthesis Methods* submitted.

Fisher DJ, Copas AJ, Tierney JF, Parmar MKB 2011. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. *Journal of Clinical Epidemiology* **64**: 949–967. DOI:10.1016/j.jclinepi.2010.11.016.

Golder S, Loke YK, Bland M 2011. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Medicine* **8**: e1001026. DOI:10.1371/journal.pmed.1001026.

Goldstein H, Browne W, Rasbash J 2002. Multilevel modelling of medical data. *Statistics in Medicine* **21**: 3291–3315. DOI:10.1002/sim.1264.

Goldstein H, Yang M, Omar R, Turner R, Thompson S 2000. Meta-analysis using multilevel models with an application to the study of class size e_ects. *Journal of the Royal Statistical Society: Series C: Applied Statistics* **49**: 399–412. DOI:10.1111/1467-9876.00200.

Guyot P, Ades AE, Ouwens MJNM, Welton NJ 2012. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology* **12**: 9. DOI:10.1186/1471-2288-12-9.

Haines TP, Hill AM 2011. Inconsistent results in meta-analyses for the prevention of falls are found between study-level data and patient-level data. *Journal of Clinical Epidemiology* **64**: 154–162. DOI:10.1016/j.jclinepi.2010.04.024.

Hall DB 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* **56**: 1030–1039.

Higgins JP, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC, Tugwell P, Wells G 2013. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. *Research Synthesis Methods* **4**: 12–25. DOI:10.1002/jrsm.1056.

Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG 2001. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* **20**: 2219–2241. DOI:10.1002/sim.918.

Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J 2001. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *Journal of the American Medical Association* **286**: 821–830. DOI:10.1001/jama.286.7.821.

Jackson C, Best N, Richardson S 2006. Improving ecological inference using individual-level data. *Statistics in Medicine* **25**: 2136–2159. DOI:10.1002/sim.2370.

Jackson C, Best N, Richardson S 2008. Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **171**: 159–178. DOI:10.1111/j.1467-985X.2007.00500.x.

Jansen JP 2012. Network meta-analysis of individual and aggregate level data. *Research Synthesis Methods* **3**: 177–190. DOI:10.1002/jrsm.1048.

Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, Salanti G 2014. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. *Value in Health* **17**: 157–173. DOI:10.1016/j.jval.2014.01.004.

Jolani S, Debray TPA, Koffijberg H, van Buuren S, Moons KGM 2015. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Statistics in Medicine* **34**: 1841–1863. DOI:10.1002/sim.6451.

Jones AP, Riley RD, Williamson PR, Whitehead A 2009. Meta-analysis of individual patient data versus aggregate data from longitudinal clinical trials. *Clinical Trials* **6**: 16–27. DOI:10.1177/1740774508100984.

Kaizar EE 2011. Estimating treatment effect via simple cross design synthesis. *Statistics in Medicine* **30**: 2986–3009. DOI:10.1002/sim.4339.

Katsahian S, Latouche A, Mary JY, Chevret S, Porcher R 2008. Practical methodology of meta-analysis of individual patient data using a survival outcome. *Contemporary Clinical Trials* **29**: 220–230. DOI:10.1016/j.cct.2007.08.002.

Kim S, Chen MH, Ibrahim JG, Shah AK, Lin J 2013. Bayesian inference for multivariate meta-analysis Box-Cox transformation models for individual patient data with applications to evaluation of cholesterol-lowering drugs. *Statistics in Medicine* **32**: 39723990.

Koopman L, van der Heijden GJMG, Glasziou PP, Grobbee DE, Rovers MM 2007. A systematic review of analytical methods used to study subgroups in (individual patient data) meta-analyses. *Journal of Clinical Epidemiology* **60**: 1002–1009. DOI:10.1016/j.jclinepi.2007.01.018.

Koopman L, van der Heijden GJMG, Grobbee DE, Rovers MM 2008a. Comparison of methods of handling missing data in individual patient data meta-analyses: an empirical example on antibiotics in children with acute otitis media. *American Journal of Epidemiology* **167**: 540–545. DOI:10.1093/aje/kwm341.

Koopman L, van der Heijden GJMG, Hoes AW, Grobbee DE, Rovers MM 2008b. Empirical comparison of subgroup effects in conventional and individual patient data meta-analyses. *International Journal of Technological Assessment in Health Care* **24**: 358–361. DOI:10.1017/S0266462308080471.

Lambert PC, Sutton AJ, Abrams KR, Jones DR 2002. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* **55**: 86–94. DOI:10.1016/S0895-4356(01)00414-0.

Larose DT, Dey DK 1997. Grouped random effects models for Bayesian meta-analysis. *Statistics in Medicine* **16**: 1817–1829. DOI:10.1002/(SICI)1097-0258(19970830)16:16<1817::AID-SIM621>3.0.CO;2-N.

Lau J, Ioannidis JP, Schmid CH 1998. Summing up evidence: one answer is not always enough. *Lancet* **351**: 123–127. DOI:10.1016/S0140-6736(97)08468-7.

Lee AH, Wang K, Scott JA, Yau KKW, McLachlan GJ 2006. Multi-level zero-innated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research* **15**: 47–61.

Li B, Lingsma HF, Steyerberg EW, Lesaffre E 2011. Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. *BMC Medical Research Methodology* **11**: 77. DOI:10.1186/1471-2288-11-77.

Li Z, Begg B 1994. Random effects models for combining results from controlled and uncontrolled studies in a meta-analysis. *Journal of the American Statistical Association* **89**: 1523–1527. DOI:10.2307/2291015.

Li Z, Meredith MP 2003. Exploring the relationship between surrogates and clinical outcomes: analysis of individual patient data vs. meta-regression on group-level summary statistics. *Journal of Biopharmaceutical Statistics* **13**: 777–792. DOI:10.1081/BIP-120024209.

Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, Hutton JL 2001. Issues in methodological research: perspectives from researchers and commissioners. *Health Technology Assessment* **5**: 1–57.

Lyman GH, Kuderer NM 2005. The strengths and limitations of meta-analyses based on aggregate data. *BMC Medical Research Methodology* **5**: 14. DOI:10.1186/1471-2288-5-14.

Mathew T, Nordström K 2010. Comparison of one-step and two-step meta-analysis models using individual patient data. *Biometrical Journal* **52**: 271–287. DOI:10.1002/bimj.200900143.

Michiels S, Baujat B, Mahé C, Sargent D, Pignon J 2005. Random effects survival models gave a better understanding of heterogeneity in individual patient data meta-analyses. *Journal of Clinical Epidemiology* **58**: 238–245. DOI:10.1016/j.jclinepi.2004.08.013.

Peters JL, Rushton L, Sutton AJ, Jones DR, Abrams KR, Mugglestone MA 2005. Bayesian methods for the cross-design synthesis of epidemiological and toxicological evidence. *Journal of the Royal Statistical Society: Series C: Applied Statistics* **54**: 159–172. DOI:10.1111/j.1467-9876.2005.00476.x.

Piepho HP, Williams ER, Madden LV 2012. The use of two-way linear mixed models in multitreatment meta-analysis. *Biometrics* **68**: 1269–1277. DOI:10.1111/j.1541-0420.2012.01786.x.

Poppe K, Doughty R, Yu C, Quintana M, Møller J, Klein A, Gamble G, Dini F, Whalley G, MeRGE collaborators 2011. Understanding differences in results from literature-based and individual patient meta-analyses: an example from meta-analyses of observational data. *International Journal of Cardiology* **148**: 209–213. DOI:10.1016/j.ijcard.2009.09.566.

Prevost TC, Abrams KR, Jones DR 2000. Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening. *Statistics in Medicine* **19**: 3359–3376. DOI:10.1002/1097-0258(20001230)19:24<3359::AID-SIM710>3.0.CO;2-N.

Ravva P, Karlsson MO, French JL 2014. A linearization approach for the model-based analysis of combined aggregate and individual patient data. *Statistics in Medicine* **33**: 1460–1476. DOI:10.1002/sim.6045.

Reeves BC, Higgins JPT, Ramsay C, Shea B, Tugwell P, Wells GA 2003. An introduction to methodological issues when including non-randomised studies in systematic reviews on the effects of interventions. *Research Synthesis Methods* **4**: 1–11. DOI:10.1002/jrsm.1068.

Resche-Rigon M, White IR, Bartlett JW, Peters SAE, Thompson SG 2013. Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine* **32**: 4890–4905. DOI:10.1002/sim.5894.

Riley RD, Kauser I, Bland M, Thijs L, Staessen JA, Wang J, Gueyffier F, Deeks JJ 2013. Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. *Statistics in Medicine* **32**: 2747–2766. DOI:10.1002/sim.5726.

Riley RD, Lambert PC, Abo-Zaid G 2010. Meta-analysis of individual participant data: rationale, conduct, and reporting. *British Medical Journal* **340**: c221. DOI:10.1136/bmj.c221.

Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Boutitie F 2008. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine* **27**: 1870–1893. DOI:10.1002/sim.3165.

Riley RD, Simmonds MC, Look MP 2007. Evidence synthesis combining individual patient data and aggregate data: a systematic review identi_ed current practice and possible methods. *Journal of Clinical Epidemiology* **60**: 431–439. DOI:10.1016/j.jclinepi.2006.09.009.

Riley RD, Steyerberg EW 2010. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods* **1**: 2–19. DOI:10.1002/jrsm.4.

Rondeau V, Michiels S, Liquet B, Pignon JP 2008. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine* **27**: 1894–1910. DOI:10.1002/sim.3161.

Royston P, Parmar MKB 2002. Flexible parametric proportional-hazards and proportional odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* **21**: 2175–2197. DOI:10.1002/sim.1203.

Royston P, Sauerbrei W 2004. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. *Statistics in Medicine* **23**: 2509–2525. PMID: 15287081.

Rubin DB. 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley and Sons, New York.

Saramago P, Sutton AJ, Cooper NJ, Manca A 2012. Mixed treatment comparisons using aggregate and individual participant level data. *Statistics in Medicine* **31**: 3516–3536. DOI:10.1002/sim.5442.

Sargent DJ 1998. A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics* **54**: 1486–1497.

Schafer JL 1999. Multiple imputation: a primer. *Statistical Methods in Medical Research* **8**: 3–15. DOI:10.1177/096228029900800102.

Schmid CH, Stark PC, Berlin JA, Landais P, Lau J 2004. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *Journal of Clinical Epidemiology* **57**: 683–697. DOI:10.1016/j.jclinepi.2003.12.001.

Schmidt AF, Rovers MM, Klungel OH, Hoes AW, Knol MJ, Nielen M, de Boer A, Groenwold RHH 2013. Differences in interaction and subgroup-specific effects were observed between randomized and nonrandomized studies in three empirical examples. *Journal of Clinical Epidemiology* **66**: 599–607. DOI:10.1016/j.jclinepi.2012.08.008.

Siannis F, Barrett J, Farewell V, Tierney J 2010. One-stage parametric meta-analysis of time-to-event outcomes. *Statistics in Medicine* **29**: 3030–3045. DOI:10.1002/sim.4086.

Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, Betts KA, Wu EQ 2012. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value in Health* **15**: 940–947. DOI:10.1016/j.jval.2012.05.004.

Signorovitch JE, Wu EQ, Yu AP, Gerrits CM, Kantor E, Bao Y, Gupta SR, Mulani PM 2010. Comparative effectiveness without head-to-head trials: a method for matching adjusted indirect comparisons applied to psoriasis

treatment with adalimumab or etanercept. *PharmacoEconomics* **28**: 935–945. DOI:10.2165/11538370-000000000-00000.

Simmonds MC, Higgins JPT 2007. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine* **26**: 2982–2999. DOI:10.1002/sim.2768.

Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG 2005. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* **2**: 209–217. DOI:10.1191/1740774505cn087oa.

Steinberg KK, Smith SJ, Stroup DF, Olkin I, Lee NC, Williamson GD, Thacker SB 1997. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. *American Journal of Epidemiology* **145**: 917–925.

Stewart GB, Altman DD, Askie L, Duley L, Simmonds M, Stewart LA 2012. Statistical analysis of individual participant data meta-analyses: a comparison of methods and recommendations for practice. *PLoS One* **7**: e46042. DOI:10.1371/journal.pone.0046042.

Stewart LA, Tierney JF 2002. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the Health Professions* **25**: 76–97. DOI:10.1177/0163278702025001006.

Stijnen T, Hamza TH, Özdemir P 2010. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine* **29**: 3046–3067. DOI:10.1002/sim.4040.

Stuart EA, Cole SR, Bradshaw CP, Leaf PJ 2011. The use of propensity scores to assess the generalizability of results from randomized trials: use of propensity scores to assess generalizability. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **174**: 369–386. DOI:10.1111/j.1467-985X.2010.00673.x.

Sud S, Douketis J 2009. The devil is in the details…or not? A primer on individual patient data meta-analysis. *Evidence-Based Medicine* **14**: 100–101. DOI:10.1136/ebm.14.4.100.

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F 1998. Systematic reviews of trials and other studies. *Health Technology Assessment* **2**: 1–276. DOI:10.3310/hta2190.

Sutton AJ, Cooper NJ, Jones DR 2009. Evidence synthesis as the key to more coherent and efficient research. *BMC Medical Research Methodology* **9**: 29. DOI:10.1186/1471-2288-9-29.

Sutton AJ, Higgins JPT 2008. Recent developments in meta-analysis. *Statistics in Medicine* **27**: 625–650. DOI:10.1002/sim.2934.

Sutton AJ, Kendrick D, Coupland C 2008. Meta-analysis of individual- and aggregate-level data. *Statistics in Medicine* **27**: 651–669. DOI:10.1002/sim.2916.

Teramukai S, Matsuyama Y, Mizuno S, Sakamoto J 2004. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. *Japanese Journal of Clinical Oncology* **34**: 717–721. DOI:10.1093/jjco/hyh138.

Thomas D, Radji S, Benedetti A 2014. Systematic review of methods for individual patient data meta-analysis with binary outcomes. *BMC Medical Research Methodology* **14**: 79. DOI:10.1186/1471-2288-14-79.

Thompson A 2009. Thinking big: large-scale collaborative research in observational epidemiology. *European Journal of Epidemiology* **24**: 727–731. DOI:10.1007/s10654-009-9412-1.

Thompson S, Kaptoge S, White I, Wood A, Perry P, Danesh J, The Emerging Risk Factors Collaboration 2010. Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies. *International Journal of Epidemiology* **39**: 1345–1359. DOI:10.1093/ije/dyq063.

Thompson SG, Turner RM, Warn DE 2001. Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical Methods in Medical Research* **10**: 375–392. DOI:10.1177/096228020101000602.

Tierney JF, Stewart LA 2005. Investigating patient exclusion bias in meta-analysis. *International Journal of Epidemiology* **34**: 79–87. DOI:10.1093/ije/dyh300.

Tobías A, Saez M, Kogevinas M 2004. Meta-analysis of results and individual patient data in epidemiological studies. *Journal of Modern Applied Statistical Methods* **3**: 176–185.

Trikalinos TA, Ioannidis JP 2001. Predictive modeling and heterogeneity of baseline risk in meta-analysis of individual patient data. *Journal of Clinical Epidemiology* **54**: 245–252. DOI:10.1016/S0895-4356(00)00311-5.

Tudur Smith C, Williamson PR 2007. A comparison of methods for fixed effects meta-analysis of individual patient data with time to event outcomes. *Clinical Trials* **4**: 621–630. DOI:10.1177/1740774507085276.

Tudur Smith C, Williamson PR, Marson AG 2005a. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Statistics in Medicine* **24**: 1307–1319. DOI:10.1002/sim.2050.

Tudur Smith C, Williamson PR, Marson AG 2005b. An overview of methods and empirical comparison of aggregate data and individual patient data results for investigating heterogeneity in meta-analysis of time-to-event outcomes. *Journal of Evaluation in Clinical Practice* **11**: 468–478. DOI:10.1111/j.1365-2753.2005.00559.x.

Turner RM, Omar RZ, Yang M, Goldstein H, Thompson SG 2000. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in Medicine* **19**: 3417–3432. DOI:10.1002/1097-0258(20001230)19:24<3417::AID-SIM614>3.0.CO;2-L.

Vaida F, Xu R 2000. Proportional hazards model with random effects. *Statistics in Medicine* **19**: 3309–3324. DOI:10.1002/1097-0258(20001230)19:24<3309::AID-SIM825>3.0.CO;2-9.

Valentine JC, Thompson SG 2013. Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions. *Research Synthesis Methods* **4**: 26–35. DOI:10.1002/jrsm.1064.

Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Hobbs B, Ibrahim JG, Kinnersley N, Lindborg S, Micallef S, Roychoudhury S, Thompson L 2014. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics* **13**: 41–54. DOI:10.1002/pst.1589.

Wakefield J, Salway R 2001. A statistical framework for ecological and aggregate studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **164**: 119–137.

van Walraven C 2010. Individual patient meta-analysis – rewards and challenges. *Journal of Clinical Epidemiology* **63**: 235–237. DOI:10.1016/j.jclinepi.2009.04.001.

Wells GA, Shea B, Higgins JP, Sterne J, Tugwell P, Reeves BC 2013. Checklists of methodological issues for review authors to consider when including non-randomized studies in systematic reviews. *Research Synthesis Methods* **4**: 63–77. DOI:10.1002/jrsm.1077.

White IR, Welton NJ, Wood AM, Ades AE, Higgins JPT 2008. Allowing for uncertainty due to missing data in meta-analysis-part 2: hierarchical models. *Statistics in Medicine* **27**: 728–745. DOI:10.1002/sim.3007.

Whitehead A, Omar RZ, Higgins JP, Savaluny E, Turner RM, Thompson SG 2001. Meta-analysis of ordinal outcomes using individual patient data. *Statistics in Medicine* **20**: 2243–2260. DOI:10.1002/sim.919.

Yamaguchi T, Ohashi Y 1999. Investigating centre effects in a multi-centre clinical trial of superficial bladder cancer. *Statistics in Medicine* **18**: 1961–1971.

Yamaguchi T, Ohashi Y, Matsuyama Y 2002. Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Statistical Methods in Medical Research* **11**: 221–236. DOI:10.1191/0962280202sm284ra.

Yau KK, Lee AH 2001. Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme. *Statistics in Medicine* **20**: 2907–2920. DOI:10.1002/sim.860.

Yau KK, McGilchrist CA 1998. ML and REML estimation in survival analysis with time dependent correlated frailty. *Statistics in Medicine* **17**: 1201–1213. DOI:10.1002/(SICI)1097-0258(19980615)17:11<1201::AID-SIM845>3.0.CO;2-7.

Yucel RM 2008. Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* **366**: 2389–2403. DOI:10.1098/rsta.2008.0038.

Yucel RM 2011. Random-covariances and mixed-effects models for imputing multivariate multilevel continuous data. *Statistical Modelling* **11**: 351–370. DOI:10.1177/1471082X1001100404.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

